DOCUMENT RESUME

ED 425 283 CE 077 493

AUTHOR	Millsap, Roger
TITLE	Self-Directed Workplace Literacy Distance Learning for Developmental Disabilities Workers: External Summative Evaluation Design.
INSTITUTION	City Univ. of New York, NY. Center for Advanced Study in Education.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, DC. National Workplace Literacy Program.
PUB DATE	1997-12-31
NOTE	22p.; For related documents, see CE 077 492-494.
CONTRACT	V198A40298-95
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Adult Basic Education; *Caregivers; *Developmental Disabilities; *Distance Education; Evaluation Criteria; Labor Education; Literacy Education; Models; Partnerships in Education; Program Effectiveness; Reading Comprehension; Reading Skills; *Summative Evaluation; Unions; *Workplace Literacy; Writing Skills
IDENTIFIERS	*Developmental Aides

ABSTRACT

This report presents an external summative evaluation plan for the Self-Directed Workplace Distance Learning for Developmental Disabilities Workers Project, a partnership between the Center for Advanced Study of Education and the Civil Service Employees Association, Inc., with the New York State Office of Mental Retardation and Developmental Disabilities. Project goals are described as documenting changes in literacy skills and other relevant variables over the course of training and linking these changes to the training intervention. Section 1 of the report describes in detail the following elements of the research design to be used: the participants, direct care workers employed in state-operated developmental disabilities centers, and overall design of data collection, a pretest-posttest nonequivalent control group design with participants in training receiving 96 hours of instruction over 6 months. Section 2 describes the measures to be used to assess expected outcomes of training: background measures, participant pretest and posttest measures, and supervisor ratings of participants at pretest and posttest. These measures include supervisor ratings of participants' job task completion and amount of improvement; participant self-ratings; reading comprehension test; writing measure; problem-solving test; and self-efficacy scale. Section 3 describes data analysis methods, including scale development, descriptive statistics, and evaluation of treatment effects using analysis of covariance. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 425-283

Self-Directed Workplace Literacy Distance Learning for Developmental Disabilties Workers: External Summative Evaluation Design

Roger Millsap, Ph. D.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Center for Advanced Study in Education
The Graduate School and University Center
of the City University of New York

CE077493

BEST COPY AVAILABLE

Self-Directed Workplace Literacy Distance Learning for Developmental Disabilties Workers: External Summative Evaluation Design

Roger Millsap, Ph. D.

Department of Psychology
Baruch College, CUNY
17 Lexington Ave
New York, NY 10010
(212) 387-1551

U.S. Department of Education Grant Award #V198A 40298-95
Project Period 01/01/95-12/31/97

Introduction

This is an external summative evaluation plan for the Self-Directed Workplace Distance Learning for Developmental Disabilities Workers Project, a partnership between The Center for Advanced Study of Education (CASE) of the City University of New York Graduate School and the Civil Service Employees Association, Inc. (CSEA), with the New York State Office of Mental Retardation and Developmental Disabilities (OMRDD) and the Government Office of Employee Relations (GOER) as helping organizations. The goals of the research component of the Distance-Learning Project are twofold. First, we seek to document changes in literacy skills and other relevant variables over the course of training. Secondly, we seek to link these changes to the training intervention. In other words, we will attempt to establish a causal relationship between the literacy training and the changes in literacy skills and other relevant variables. Both goals require that the expected outcomes of training be carefully measured. The measures developed for this purpose will be described below. More generally, the research design to be used will be described in detail, along with the general data analytic methods to be used once the data become available.

This report is presented in three sections. The first section gives a general description of the research design. The following section describes the measures to be used in the study. The third section describes the methods of data analysis to be

applied to the data that will emerge from the study.

Research Design

Participants

The participants in the Distance Learning Project are direct care workers employed in state-operated developmental disabilities centers in New York State. Participation in the Project is voluntary. All participants will receive 100% released-time for participation. It is expected that approximately 380 employees will receive training across the three years of the Project, with about 95 participants in the first year, 186 in the second year, and 99 in the third year.

The participants work in 80 different work sites around the state, organized within five Developmental Disabilities Services Offices (DDSOs).[1] About 75% of the participants are Developmental Aides, with the remaining 25% who work under a variety of job titles, but aspire to the Developmental Aide title. More than half of the participants work in community homes. The remaining participants work in developmental centers and seek to make the transition to community homes.

Participants range in age from 25 to 60, with the majority between 35 and 45 years of age.

In addition to participants who receive training at a given

---

[1] After the beginning of the project period, there was a legislative mandate to administratively consolidate the DDSO's, reducing the number of DDSO's served by this project from seven to five. However, it is important to note that the same territory described in the grant application is covered now. A consolidated DDSO serves the same geographic area and number of persons previously covered by two separate entities.

time, groups of participants who are scheduled to recieve training at a later date will be designated as controls for purposes of data collection. Hence the control-group members will be participants who are wait-listed. If there are not enough participants on the waiting lists, additional controls who are also Developmental Aides will be found. As outlined below, the measures collected from control-group members will be almost identical to those collected from individuals who are receiving training, and will be collected at the same time as treatment group data. Data on controls will be collected during month 7-18 of the project period. It is expected that the total number of control individuals will be about one fourth of the total number of participants, or about 100 individuals.

All participants will be asked to sign human subjects consent forms in compliance with CUNY Graduate School procedures.

Design

The overall design of the data collection is best described as a pretest-posttest nonequivalent control group design (Cook and Campbell, 1979). Those who receive literacy training at a given time are considered to be members of the "treatment" group for that time period, while those who are wait-listed and are designated for data collection as control members are considered to be members of the "control" group. Although members of both groups will be drawn from the same pool of employees, the allocation of employees to treatment vs. control conditions will not be strictly random. This fact allows for nonequivalence

between the two groups, or the possibility of pre-existing differences between the groups.

Both control and treatment group members will be pretested and posttested using measures to be described below. Pretesting for trainees will take place during the initial entrance into training. Posttests will be given upon completion of training (after six months). The time interval separating pretests and posttests for control group members will also be six months. Whenever possible, the posttest measurements for control group members will also serve as pretest measurements upon their entrance into training. The only exceptions to this rule will involve cases in which a fairly long time interval separates the posttest and entrance into training. The number of individuals in this category is expected to be small.

Participants in training will receive 96 hours of instruction over a six month period. Instruction will occur for four hours per week for 24 weeks. This 24 week period is divided into four quarters of six weeks each, as noted in the grant proposal. The first quarter consists of the "core" instructional phase that is essentially identical for all participants. The remaining three quarters are based in the Individualized Educational Plans (IEPs) developed by individual participants in cooperation with their instructors. Hence the nature of the instruction received in the last three quarters will vary among participants. Once the data are collected, it may be possible to group participants into finer classifications based on common

features of their IEPs. These groups may then be contrasted in order to study differences in outcomes as a function of type of instruction.

During the final three months of the study period, a group of 10 participants who received training during the first year of training will again be interviewed regarding job performance and career advancement. These 10 individuals will be selected to have adequate variability on the pretest measures. The information gained from these individuals will be used to explore the long-term changes associated with participation in the Distance Learning project.

As is true in any field study, it is expected that some percentage of the participants will drop out of the study before both pretest and posttest information can be collected. Whenever possible, an effort will be made in each case to discover the reason for the dropout. Assuming that complete demographic and pretest information will be available for all individuals who are dropouts, it will be possible to compare dropouts with the remaining participants to discover important differences between the two groups. These comparisons will help to determine whether the loss of participants is likely to distort the analysis of treatment effects (Little and Rubin, 1987).

Measures

Background Measures.

Background information will be recorded for each individual that enters the project. This information includes the

individual's age, gender, marital status, number of dependents, job title, job location, job tenure, whether full or part-time, the language used in childhood, the language used most often now, the highest grade completed in school, whether the individual has had any non-credit courses in reading or writing, or any other training that the individual has received. All of this information is gathered when the participant first enters the project, for both treatment and control individuals.

Pretest and Posttest Measures.

Pretest measures will be given to both treatment and control individuals during their initial entrance into the program. Posttest measures will be given at the end of the six month training period. The time required for either set of measures to be completed by a participant is about 1 1/2-2 hours. A proctor (usually a supervisor or other local team member) will be present during the testing of each participant to ensure the timely completion of the measures. The pretest and posttest measures are identical except as noted below.

In addition to the measures to be completed by the participants, the supervisor of each participant will be asked to complete some ratings of the participant at both pretest and posttest. These ratings are collected on the same time schedule as the other pretest and posttest measures. The ratings consist of 12 questions regarding how well the participant is able to complete various job tasks. The tasks involve basic reading, writing, and arithmetic skills. Ratings are given on a four-

point scale, with an additional rating category for "not applicable" to be used if the participant never does the task in question. At posttest, several additional ratings are requested regarding the degree to which the participant's reading, writing, math, and problem-solving skills have improved since pretest. Ratings are again given on a four-point scale. These pretest and posttest ratings will be denoted the "Supervisor Ratings" in what follows.

Along with the Supervisor Ratings, each participant is asked to rate his or her own task completion using the same set of 12 job tasks that were rated by their supervisor. The response scale is the same as that used by their supervisor. At posttest, the participant is also asked to rate the amount of improvement on the same four skills rated by their supervisor. These participant self-ratings will be denoted the "Participant Self-Ratings" in what follows.

All measures in this evaluation are customized to the job, and are locally developed. Each participant will complete a test of reading comprehension at pretest and posttest, denoted the "Reading Comprehension Test" in what follows. Examinees are given three reading selections, with each selection followed by a set of multiple-choice questions that refer to the selection. The reading selections contain job-related material similar to that encountered on the job. A total of 10 multiple-choice questions are given. Each examinee is given 20 minutes to complete the test. The examinee's score is calculated as an

unweighted total across these 10 questions.

The next test is a direct writing assessment measure, denoted the "Writing Measure" here. In this test, examinees are asked to write a description of their job as if they were describing the job to a coworker. Each examinee is given 20 minutes to write the essay. The essays will be scored by two readers. These readers will be given training in the scoring task prior to any grading. The scoring system is an analytic system that rates the essay on five dimensions: content, organization, vocabulary, language use, and mechanics. Each dimension is scored on a four-point scale ranging from "very poor" to "excellent to very good". The total number of points given for each essay will therefore range from 5 to 20. The two readers are instructed to rate an essay independently, and then to reach consensus if their ratings differ. If a consensus cannot be reached, a third reader will read the essay. The final score for the essay will be an average of the three reader's ratings.

Each participant will then complete a test of problem-solving skills, denoted the "Problem-Solving Test" here. In this test, the examinee is presented with a series of job-related scenarios. Each scenario consists of a problem situation typical of those that might be encountered on the job. The examinee must write a paragraph describing what he or she feels is the best way of resolving the problem. Four scenarios are presented in the test, and each requires a separate written response. The written

responses will be scored by readers who have been trained for this scoring task. "Best case" solutions to each scenario (and the scenarios themselves) have been developed in collaboration with present and former staff members of the New York State Office of Mental Retardation and Developmental Disabilities (OMRDD). The response to each scenario is scored on a four-point scale to indicate its similarity to the best case solution. Total scores across the four scenarios range from 4 to 16 points.

The final measure used in the pretest and posttest is a self-efficacy scale that concerns job-related competencies. This measure will be denoted the "Self-Efficacy Scale" here. In this scale, the examinee is presented with five work situations in which a general task is to be completed. Within each situation, the examinee is asked about the degree to which he or she is sure that various activities could be successfully completed. All of the activities concern different subtasks that must be performed in the situation. These subtasks involve the three different basic literacy skills of reading, writing, and math. The examinee responds to each question on a seven-point scale ranging from "Not at all sure" to "Very sure". A total of 18 questions are asked across the five situations. Total scores may be calculated as sums of item scores across all five situations, or subtest scores can be calculated to correspond to specific literacy skills.

Additional Measures.

There are several additional measures that will be taken

during the six month training period which are not part of the pretest/posttest set. First, job attendance and absenteeism data for all participants (treatment and controls) will be made available by the employers. These data will be available for the entire six months of the project, permitting any trends in the data to be studied. Secondly, different measures of the participant's utilization of training services will be available. For example, the frequency of e-mail usage will be available for participants who work in DDSO's that support e-mail. Also, frequency of participant's telephone communication with instructors will be available for all DDSO's. Neither of these measures will be available for individuals in the control condition however, for obvious reasons.

A third variable is the type of training received by a participant in the second, third, and fourth quarters of the training period. Recall that following the first quarter of core instruction, participants design their own training for the remaining three quarters in cooperation with their instructors. The nature of this training will vary among participants, creating a new variable "type of training" that can itself serve as a way of grouping participants for later comparisons. The precise definition of this variable must await completion of at least one six-month training period in order to acquire data on the variety of training types.

Data Analysis

Scale Development

Study measures that are multiple-item scales (Supervisor Ratings, Participant Self-Ratings, Reading Comprehension test, Problem-Solving Test, and the Self-Efficacy scale) will first be evaluated for internal consistency. Internal consistency will be evaluated using both alpha coefficients and factor analysis. Exploratory factor analysis will be used to check for the existence of multiple factors. A check for the fit of a single-factor model in any scale can be made using confirmatory factor analysis software such as LISREL (Joreskog and Sorbom, 1989). If multiple factors emerge that are interpretable, the possibility of multiple scales will be considered. For example, the Self-Efficacy scale may yield multiple scales concerning specific competencies, rather than a single general efficacy scale. In some cases, it may be necessary to drop items that show poor item statistics (e.g., low variance). Alpha coefficients will be used to measure the internal consistency of any final scales.

All scale development will be done separately for pretest and posttest measures. It is possible that a given scale will not maintain an identical structure from pretest to posttest. Changes in structure can, in some cases, result from the intervention (Millsap and Hartog, 1988). For any given scale, it is possible to compare the treatment and control groups at pretest or posttest on the factor structure using confirmatory factor analysis. This option may be pursued if preliminary

analyses indicate group differences in structure, or marked changes in structure from pretest to posttest.

Correlations between Supervisor and Participant Self-Ratings will be calculated at both pretest and posttest. Correlations can be calculated at the item level and at the level of the total scale. These correlations will indicate the level of agreement between the two sources.

Data on the interrater agreement in the grading of the Writing Measure will be available apart from the data collection.

Descriptive Statistics

Descriptive statistics will be calculated on all outcome measures and demographic variables. Statistics will be calculated separately for pretest and posttest measures, with correlations between pretest and posttest measures also being calculated. Breakdowns of these statistics by treatment vs. control group will be performed.

One important goal in this analysis will be to document any differences between the control and treatment groups at pretest. Differences may exist either in demographic variables or in pretest scores. Careful documentation of such differences is important in helping to establish the equivalence, or lack of equivalence, between control and treatment groups.

Another grouping that will be important in the descriptive phase concerns individuals with both pretest and posttest data, and individuals who did not remain in the program for the posttest. The latter group are the "dropouts". It is important

to document any differences between these two groups, both in terms of demographics and on the pretest measures. It may be possible to further classify the dropouts according to their reason for leaving the study (e.g., voluntary vs. involuntary). This possibility can be explored if the number of dropouts is substantial.

Evaluation of Treatment Effects

The basic tool in the evaluation of possible treatment effects in this study will be the analysis of covariance (ANCOVA) using the pretest measures as covariates. In this analysis, treatment and control individuals are first matched statistically in terms of pretest scores, followed by comparisons of group mean differences on the posttest scores within these matched sets. The matching on the pretest scores attempts to eliminate any group differences on the posttest scores that may be attributable to differences at pretest. The ANCOVA can, under fairly general conditions, lead to a more powerful analysis than simple group comparisons of gain scores (Cook and Campbell, 1979). Note that the ANCOVA's will be limited to individuals who have both pretest and posttest data.

The ANCOVA's will be performed separately for the various pretest and posttest measures since the number of such measures is not large. If multiple scales emerge for some of these measures, multivariate ANCOVA's may be considered. One preliminary assumption required for meaningful interpretation of the ANCOVA is homogeneity of the regression slopes for the

regression of posttest scores on the pretest. This assumption will be checked prior to the analyses.

Effect sizes in the ANCOVA can be expressed in terms of the group differences in the posttest means after adjustment for the pretest. These adjusted means will be calculated for any comparisons that reach statistical significance.

As noted earlier, it is important to establish the equivalence of the treatment and control groups at pretest. This goal will be pursued by comparing the two groups on all pretest scales using t-tests. Comparisons will also be done using relevant demographic variables. Some of these variables are simple categorizations (e.g., gender). Treatment and control groups can be statistically compared on such measures using chi-square tests. Significant and meaningful differences between the treatment and control groups at pretest will complicate the causal interpretations of any ANCOVA results.

Some of the items in the Supervisor and Participant ratings are only given at posttest. These items ask the rater to rate the degree of change from pretest to posttest in certain job skills. Treatment and control groups can be compared directly on these posttest measures using simple t-tests.

Absenteeism and attendance data will be available for the entire six months of the training period, for both treatment and control individuals. These data will be analysed by first aggregating the data within each individual on a monthly basis (i.e., monthly attendance and absenteeism figures) for the six

months of the training period. Each individual will have six scores after aggregation. It is hoped that attendance will show an increasing trend, and absenteeism a decreasing trend, across the six months of the project. This hypothesis will be evaluated using a 2x6 repeated measures analysis of variance, with trends across the 6 months being contrasted between treatment and control groups.

Measures of the treatment individuals' utilization of services will also be available, as noted earlier. Correlations between these measures and the posttest scale scores will be calculated. Substantial correlations may suggest that it will be useful to divide the treatment group into subgroups based on level of utilization. In this event, the above ANCOVA analyses may be repeated using the control group and the multiple treatment groups created by level of utilization. It is possible that treatment effects will only be apparent among individuals with higher levels of service utilization.

All of the foregoing analyses have assumed that data from different cycles of the project are pooled into a single analysis, with no attempt to examine differences in any effects according to cycle. We have no good reason to hypothesize variations in treatment effects over cycles, other than the possibility that early cycles may show different trends due to problems with the start-up. Disaggregation by cycle would also shrink the effective sample sizes for the analyses. In the event that treatment effects are not clearly demonstrated in the pooled

data, it may be useful to disaggregate and examine effects that are specific to the different cycles.

The covariates to be used in the ANCOVA's described above are the pretest scores on the scales being analysed at posttest in a given analysis. Clearly, it is possible to use additional covariates in a given analysis. For example, in the ANCOVA of the Problem-Solving Test it would be possible to use the Writing Measure as a covariate in addition to pretest scores on the Problem-Solving Test. This additional covariate may be useful because the Problem-Solving Test requires a written response, and writing skills may be influential. On the other hand, the use of additional covariates may complicate the interpretation of the results and may lead to violations of the assumptions for the ANCOVA. The best strategy may be to pursue additional covariates only if no meaningful results are found using the pretest scores themselves as covariates.

A further direction for the analysis is to investigate variations in treatment effects as a function of demographic or job-related variables, such as age or job tenure. One strategy for doing so is to incorporate these variables into the ANCOVA as grouping variables, producing a factorial ANCOVA. Continuous variables such as age may be grouped for this purpose. This analysis would permit both the "main effects" of these variables and their interactions with the treatment/control distinction to be studied. A disadvantage of such analyses is that the number of individuals per "cell" is reduced as more grouping variables

are added. An alternative approach that could be used with continuous variables such as age is to introduce them as additional covariates. This approach would require that the added variable does not interact with the treatment/control status in its effect on the posttest scale however. Also, the added covariate is simply being used for purposes of statistical control in this approach, rather than being studied for its own effects. The choice to be made here can await the outcome of the preliminary ANCOVA's described earlier.

Finally, the comparison of dropouts to individuals who stayed in the program can proceed by comparing the groups on both demographic variables and pretest scores using t-tests. Chi-square analyses can be used for categorical demographic variables. As noted earlier, it may be useful to divide the dropouts into subgroups depending on their stated reason for leaving. This option will be pursued only if there are sufficient numbers of dropouts to warrant the approach. Dropout rates between treatment and control conditions will also be compared. Significant differences here may complicate the interpretation of the ANCOVA results if dropouts tend to differ on the pretests from individuals who stayed in the program.

Conclusion

The above plan for the data analysis will allow for a full description of the differences between the treatment and control groups in changes in literacy skills, meeting the first goal of the research component of the project. The second goal, that of

demonstrating that group differences arose as effects of the treatment, is more difficult to achieve. The central difficulty is the lack of randomization in the assignment of individuals to treatment and control conditions. The extent of the nonequivalence that will result from this assignment will only be partially known once all of the data are in, as the groups may differ in ways not revealed in the measured variables. Hence there are inherent limitations in the nonequivalent control-group design as a tool for causal inference. Within these limitations however, the planned analyses should be optimal in eliminating alternative explanations for group differences in outcomes.

References

Cook, T.D. & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand-McNally.

Joreskog, K.G. & Sorbom, D. (1989). LISREL VII user's reference guide. Mooresville, Indiana: Scientific Software.

Little, R.J.A. & Rubin, D.B. (1987). Statistical analysis with missing data. New York: John Wiley & Sons.

Millsap, R.E. & Hartog, S.B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. Journal of Applied Psychology, 73, 574-584.